EM for GMM

Parameters $\theta = (\pi_1, \dots, \pi_C, \theta_1, \dots, \theta_C)$ where $\theta_c = (\mu_c, \sigma_c^2)$.

Current parameters θ^i.

Expected complete log-likelihood

$$\sum_n E_{p(z_n \mid x_n, \theta^i)}\left[\ell(\theta; z_n, x_n)\right]$$

$$= \sum E\left[\log \prod_c \left(\pi_c \, p(x_n \mid z_n = c, \theta_c)\right)^{I(z_n = c)}\right]$$

$$= \sum E\left[\sum_c I(z_n = c) \log \pi_c + \sum_c I(z_n = c) \log p(x_n \mid \theta_c)\right]$$

$$= \sum_{n,c} \underbrace{E[I(z_n = c)]}_{= p(z_n = c \mid x_n, \theta^i) = r_{nc}} \log \pi_c + \sum_{n,c} \underbrace{E[I(z_n = c)]}_{= r_{nc}} \log p(x_n \mid \theta_c)$$

$$= \overbrace{\sum_c \left[\sum_n r_{nc}\right] \log \pi_c}^{(A)} + \sum_c \overbrace{\sum_n r_{nc} \log p(x_n \mid \theta_c)}^{(B_c)}$$

n

(A) and each (B_c) can be maximised separately.

(A) max. by

def as $\sum_n r_{nc}$

$$\pi_c = \sum_n r_{nc} / N = r_c / N$$

For (B_c),

$$\text{let} \quad \alpha_c = \frac{1}{\sigma_c}$$

and

$$f_c(\mu_c, \alpha_c) = \sum_n r_{nc} \log\left[\frac{\alpha_c}{\sqrt{2\pi}} e^{-\frac{\alpha_c^2}{2}(x_n-\mu_c)^2}\right]$$

$$= \sum r_{nc} \log \alpha_c - r_{nc}\frac{\alpha_c^2}{2}(x_n - \mu_c)^2 + \underbrace{C}_{\text{some const.}}$$

$$\frac{df}{d\mu_c} = \sum_n r_{nc}\alpha_c^2 (x_n - \mu_c) \quad \text{and}$$

$$\frac{df}{d\mu_c} = 0 \Rightarrow \sum_n r_{nc}\cancel{\alpha_c^2} x_n = \sum_n r_{nc}\cancel{\alpha_c^2}\mu_c$$

so $\quad \mu_c = \sum_n x_n r_{nc} \Big/ r_c$

—

$$\frac{df}{d\alpha} = \sum_n \frac{r_{nc}}{\alpha_c} - r_{nc}\alpha_c (x_n - \mu_c)^2$$

$$\frac{df}{d\alpha} = 0 \Rightarrow \frac{1}{\alpha_c}\sum_n r_{nc} = \alpha_c \sum_n r_{nc}(x_n - \mu_c)^2$$

so $\quad \sigma^2 = \frac{1}{\alpha^2} = \sum_n r_{nc}(x_n - \mu_c)^2 \Big/ r_c.$

HMM

Complete data

$$D = \{ (x^n_{1:T}, z^n_{1:T+1}) \mid n \in [N] \}$$

Likelihood

$$L(A, B; D)$$

$$= \prod_{n=1}^{N} \prod_{t=1}^{T} p(x^n_t \mid z^n_t) \, p(z^n_{t+1} \mid z^n_t)$$

$$= \prod_{n} \prod_{t} \prod_{s,k} B_{sk}^{I(x^n_t = s, z^n_t = k)} \prod_{k,\ell} A_{\ell k}^{I(z_t = k, z_{t+1} = \ell)}$$

$$= \left[\prod_{s,t} \prod_{n,t} B_{sk}^{I(x^n_t = s, z^n_t = k)} \right] \left[\prod_{k\ell} \prod_{n,t} A_{\ell k}^{I(z_t = k, z_{t+1} = \ell)} \right]$$

$$= \prod_{n,t} B_{sk}^{M_{sk}} \prod_{k\ell} A_{\ell k}^{N_{\ell k}}$$

$$M_{sk} = \sum_{n,t} I(x^n_t = s, z^n_t = k)$$

$$N_{\ell k} = \sum_{n,t} I(z^n_t = k, z^n_{t+1} = \ell)$$

So MLE

$$B_{sk} = \frac{M_{sk}}{\sum_s M_{sk}} \quad \text{and} \quad A_{ek} = \frac{N_{ek}}{\sum_e N_{ek}}$$

$$\ell(A, B; D) =$$

$$\sum_{s,k} \sum_{n,t} I(x_t^n = s, z_t^n = k) \log B_{sk}$$

$$+ \sum_{k,\ell} \sum_{n,t} I(z_t^n = k, z_{t+1}^n = \ell) \log A_{\ell k}$$

EM for HMM

$$D = \{ x^n_{1:T} : n \in [N] \}$$

ECLL

$$\sum_n E_{z_{1:T+1} | A^i, B^i, x^n} \left[\ell(A, B; z_{1:T+1}, x^n_{1:T}) \right]$$

$$= \sum_n E \left[\log \prod_t \prod_{s,k} B_{sk}^{I(x^n_t = s, z_t = k)} \prod_{k,\ell} A_{\ell k}^{I(z_t = k, z_{t+1} = \ell)} \right]$$

$$= \sum_n E \left[\sum_{s,k} \sum_t I(x^n_t = s, z_t = k) \log B_{sk} + \sum_{k,\ell} \sum_t I(z_t = k, z_{t+1} = \ell) \log A_{\ell k} \right]$$

$$= \sum_{s,k} \underbrace{\left[\sum_{n,t} p(x^n_t = s, z_t = k | A^i, B^i, x^n_{1:T}) \right]}_{\hat{M}_{sk}} \log B_{sk}$$

$$+ \sum_{k,\ell} \underbrace{\left[\sum_{n,t} p(z_t = k, z_{t+1} = \ell | A^i, B^i, x^n_{1:T}) \right]}_{\hat{N}_{\ell k}} \log A_{\ell k}$$

Max by

$$B_{sk} = \frac{\hat{M}_{sk}}{\sum_s \hat{M}_{sk}} \quad \text{and} \quad A_{\ell k} = \frac{\hat{N}_{\ell k}}{\sum_\ell \hat{N}_{\ell k}}$$

EM Motivation

Let

$$\mathcal{L}(q(z), \theta) = \sum_z q(z) \log \frac{p(x, z \mid \theta)}{q(z)}$$

and

$$KL(q(z) \,\|\, p(z \mid x, \theta)) = \sum_z q(z) \log \frac{q(z)}{p(z \mid x, \theta)}$$

where θ is all parameters.

Claim

$$\log p(x \mid \theta) = \mathcal{L}(q(z), \theta) + KL(q(z) \,\|\, p(z \mid x, \theta))$$

Proof: RHS

$$= \sum_z q(z) \log \frac{p(x, z \mid \theta)}{p(z \mid x, \theta)}$$

$$= \sum_z q(z) \log p(x \mid \theta)$$

$$= \log p(x \mid \theta)$$